12. Newton's Method

- Newton's method
- Cholesky decomposition.

- Homework 2 due date Feb 9, EOD
- Midterm Feb 22

Newton's method.

$$\min_{x \in \mathbb{R}^n} f(x). \qquad f: \mathbb{R}^n \to \mathbb{R}$$

continuously diff.

- Newton's method is second order method (Hessian).
- Given an iterate x_k, the next iterate x_{k+1}

$$x_{k+1} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \underbrace{f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k)}_{\text{quadratic approx of } f \text{ at } x_k}.$$

- x_{k+1} is unique if $\nabla^2 f(x_k) \succ 0$.

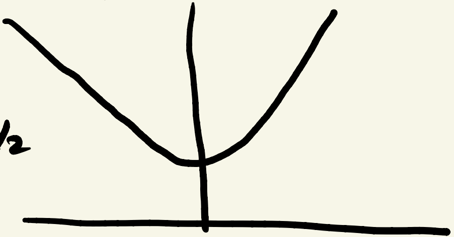
$$\nabla f(x_k) + \nabla^2 f(x_k)(x - x_k) = 0$$

$$x_{k+1} = x_k - \alpha_k (\nabla^2 f(x_k))^{-1} \nabla f(x_k).$$

- The update x_{k+1} with $\alpha_k = 1$ is called pure Newton's method.
- $-(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$ is descent direction if $\nabla^2 f(x_k) \succ 0$.

Convergence of Newton's Method.

Q. Will pure Newton's method converge if $H_k \succ 0$ for all k?



$x=0$

Example: $f(x) = \sqrt{1+x^2}$

$$f'(x) = \frac{x}{\sqrt{1+x^2}}, \quad f''(x) = (1+x^2)^{-3/2} > 0$$

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} = x_k - \frac{1}{\sqrt{1+x^2}}(1+x^2)^{3/2} = -x_k^3 = (-1)^k x_0^{3^{k+1}}$$

Therefore:

$$x_k \to \begin{cases} 0 & \text{if } |x_0| < 1 \\ \pm 1 & \text{if } |x_0| = 1 \\ \text{diverge} & \text{if } |x_0| > 1 \end{cases}$$

- Pure Newton may diverge even if $H_k \succ 0$.
- In practice, dampen ($\alpha_k < 1$).

Convergence of Newton's Method. (Thm 5.2 in Amir Beck)

Suppose $f: \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable.

and: ① $\nabla^2 f(x) \succ \varepsilon I$ for some $\varepsilon > 0$ and for all x.

② $\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq L \|x-y\|$ for all x, y and some $L > 0$.

Then, the pure Newton iteration satisfies:

$$\|x_{k+1} - x^*\|_2 \leq \frac{L}{2\varepsilon} \|x_k - x^*\|_2^2$$

In addition, if $\|x_0 - x^*\|_2 \leq \frac{\varepsilon}{L}$, then

$$\|x_{k+1} - x^*\| \leq \left(\frac{2\varepsilon}{L}\right)\left(\frac{1}{4}\right)^{2^k}, \quad k = 0, 1, 2, \dots$$

(Local quadratic convergence). $\|x_{k+1} - x_*\| \leq \frac{(\gamma - 1)}{(k+1)} \|x_k - x_*\|$

Rates of convergence

If $x_k \to x^*$ and there exist a real number $M > 0$ and $p \geq 1$ s.t.

$$\lim_{k \to \infty} \frac{\| x_{k+1} - x_* \|}{\| x_k - x_* \|^p} = M.$$

then, we say p is the rate of convergence of $\{x_k\}$.

eg: $x_k = \left(\frac{1}{10}\right)^{2^k}$ has rate of convergence of 2.

Damped Newton's Method.

$$x_{k+1} = x_k - \alpha_k \underbrace{\left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k)}_{d_k}.$$

Input: $\varepsilon > 0$ tolerance

x_0 – initial guess.

$\mu \in (0,1)$ back tracking parameter.

- Compute the Newton direction d_k by solving

$$\nabla^2 f(x_k)\, d_k = -\nabla f(x_k).$$

- set $\alpha_k = 1$. while

$$f(x_k) - f(x_k + \alpha_k d_k) \overset{>}{<} -\mu \alpha_k \nabla f(x_k)^T d_k$$

$$\alpha_k \leftarrow \alpha_k / 2$$

- $x_{k+1} = x_k + \alpha_k d_k$

- stop if $\|\nabla f(x_{k+1})\| < \varepsilon$.

Factorization

Previously, we have seen that any matrix $A \in \mathbb{R}^{m \times n}$ has QR factorization, i.e $A = QR$.

$$Ax = b \iff Rx = Q^T b.$$

If $A > 0$, other factorization are available:

- Eigenvalue decomposition: U - orthogonal, Λ - diagonal.

$$A = U \Lambda U^T, \quad Ax = b \iff \Lambda y = U^T b, \quad y = U^T x.$$

- Cholesky decomposition: $L > 0$, L - lower triangular.

$$A = LL^T, \quad Ax = b \iff \underbrace{Ly = b}_{\text{back solve}}, \quad \underbrace{L^T x = y}_{\text{forward solve}}.$$

Rosenbrock function.

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2.$$

Solution: $(x_1, x_2) = (1, 1)$.

$$\nabla^2 f(1,1) = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix}, \quad \kappa(\nabla^2 f(1,1)) \text{ Large.}$$